

L Number	Hits	Search Text	DB	Time stamp
19	10	(("6441832") or ("6195689") or ("6247052") or ("6473902") or ("5778187") or ("6185625") or ("6332147") or ("6522342") or ("6449365") or ("6222530")).PN.	USPAT	2003/11/26 15:45
20	2	((("6452609") or ("6289380"))).PN.	USPAT	2003/11/26 15:54
21	1	((("6452609") or ("6289380"))).PN.) and encod\$6	USPAT	2003/11/26 16:18
22	0	6567796.pn. and encod\$6	USPAT	2003/11/26 16:52
23	3831	administrator\$3 with user\$3	USPAT	2003/11/26 16:48
24	1	6452609.pn. and (digital\$3 file\$5)	USPAT	2003/11/26 16:48
25	2	("6452609" "6567796").pn. and (convert\$6 convers\$6 encod\$6)	USPAT	2003/11/26 16:56
26	112	(media adj stream\$5) with (encod\$5 digital)	USPAT	2003/11/26 16:55
27	0	("6452609" "6567796").pn. and (compress\$5 packetiz\$6)	USPAT	2003/11/26 16:58
28	0	("6452609" "6567796").pn. and (packet\$5 header\$5)	USPAT	2003/11/26 17:30
29	2	hypertext with (digital adj file\$3)	USPAT	2003/11/26 17:02
30	57	hypertext near8 (digital )	USPAT	2003/11/26 17:02
31	1	("6452609" "6567796").pn. and ((window media real) adj player\$3)	USPAT	2003/11/26 17:43
32	2	("6452609" "6567796").pn. and (web page\$5)	USPAT	2003/11/26 17:21
34	3	((media voice audio) adj encod\$5) with (web adj page\$5)	USPAT	2003/11/26 18:35
35	5	((media voice audio) adj encod\$5) with (web )	USPAT	2003/11/26 17:27
36	2	("6452609" "6567796").pn. and (hyperlink\$5 link\$3 url)	USPAT	2003/11/26 18:46
37	304	manual near3 encod\$5	USPAT	2003/11/26 18:11
38	5	(manual near3 encod\$5 near9 (voice audio media multimedia))	USPAT	2003/11/26 18:05
39	1	(manual near3 encod\$5 ) and 6053415.pn.	USPAT	2003/11/26 17:39
40	2	("6452609" "6567796").pn. and ( window media real player\$3)	USPAT	2003/11/26 17:44
41	2	("6452609" "6567796").pn. and ( multimedia audio video )	USPAT	2003/11/26 17:45
42	2	("6452609" "6567796").pn. and (administrat\$6 manager\$5 central\$6)	USPAT	2003/11/26 17:46
43	0	(manual near3 encod\$5) with (download\$5)	USPAT	2003/11/26 18:06
45	64	manual adj encod\$5	USPAT	2003/11/26 18:24
46	130	(web adj server\$5) with gui	USPAT	2003/11/26 18:24
47	5	((media real quicktime ) adj player\$3) with (stop start) with download\$6	USPAT	2003/11/26 18:39
48	19	((media real quicktime ) adj player\$3) with (stop start)	USPAT	2003/11/26 18:39
49	1	("6452609" "6567796").pn. and (download\$6)	USPAT	2003/11/26 18:46
-	1	("5835717").PN.	USPAT	2003/11/26 15:43
-	3	((("6434680") or ("6489979") or ("6516356"))).PN.	USPAT	2003/11/26 13:38
-	723	(graphical adj user\$3 adj interface\$5) with browser\$5	USPAT	2003/11/26 13:39
-	93	(graphical adj user\$3 adj interface\$5) with encod\$5	USPAT	2003/11/26 13:39

This is Google's cache of <http://xml.coverpages.org/barth00.html>.

Google's cache is the snapshot that we took of the page as we crawled the web.

The page may have changed since that time. Click here for the [current page](#) without highlighting.

To link to or bookmark this page, use the following url: <http://www.google.com/search?q=cache:YUUmhv6qNUJ:xml.coverpages.org/barth00.html+manual+encoding++&hl=en&start=18&ie=UTF-8>

*Google is not affiliated with the authors of this page nor responsible for its content.*

These search terms have been highlighted: **manual encoding**

[Canonical version on: <http://bcf.usc.edu/~lbarth/>]

# Electronic Edition of the Midrash Pirqe Rabbi Eliezer Creating an Encoding Manual

Lewis M. Barth

## INTRODUCTION

This paper will deal with several structural and **encoding** issues encountered in the process of creating a **manual** for **encoding** an electronic edition of *Pirqe Rabbi Eliezer* (Pirqe R. El.), *the Chapters of Rabbi Eliezer*. Pirqe R. El. is a midrashic retelling of significant aspects of the biblical narrative, from the creation story through the Book of Esther was written in the Land of Israel probably during the eighth century CE, i.e., in the early Muslim period. It contains references to Aisha, the wife of Mohammed and to Fatima, his daughter. Pirqe R. El. seems to be a kind of narrative reader's digest of rabbinic traditions on the biblical text. In fact it contains strong echoes of material known only from Old Testament Pseudepigrapha, as well as mystical and astrological material. The language of Pirqe R. El. is Hebrew with a few non-Hebrew loan words in transliteration.

Pirqe R. El. was exceedingly popular in medieval and pre-modern traditionalist Jewish literary circles. It is preserved more than twenty complete manuscripts containing fifty-two to fifty-four chapters and more than seventy-five partial manuscripts and fragments. In addition, over thirty printed editions of Pirqe R. El. have appeared since the sixteenth century. Recently, scholarly interest in Pirqe R. El. has focused on literary, historical and interpretative issues.<sup>1</sup>

There is no scholarly edition of this text in the modern sense of this term. An electronic text of Pirqe R. El. exists and commercially available.<sup>2</sup> However, the present e-text is simply the encoded version of a semi-critical eclectic edition which appeared in the 1940's. That edition, originally prepared by a scholar named Michael Higger, is based on three manuscripts whose relationship has not been fully determined. In both Higger's version and its electronic copy, there is no mark-up. This includes the fact that there are no references indicating the source of hundreds of citations from numerous biblical or rabbinic passages which are found in the text.

The initial goal of this project was to create a critical edition of Pirqe R. El. The goal has now expanded to include electronic publication of all Pirqe R. El. manuscripts and fragments in two forms: digital facsimiles and transcription with hypertext links. There are two reasons for this: 1) the quantity of textual material and 2) recent hypotheses regarding the development of medieval Hebrew manuscripts which argue that each manuscript of a work is a completely new literary creation.<sup>3</sup> Thus the need to present visually a representation of each manuscript (at least the major ones), with the possibility of comparing complete readings of specific passages on the fly -- something possible only electronically.

This paper will elaborate on some matters raised in the Introduction and concentrate on technical areas necessary for <http://www.google.com/search?q=cache:YUUmhv6qNUJ:xml.coverpages.org/barth00.html+manual+encoding++&hl=en&start=18&ie=UTF-8>

preparing an **Encoding Manual** for the project-both to remind me of what I decided to do and guide others who may become involved in **encoding** and tagging.

## SGML/TEI ELEMENT and ID ATTRIBUTE DESIGNATIONS FOR DIVISIONS of the TEXT

The first issue emerges from document analysis. It concerns the simple question: as this is a "prose" document, how should the text of Pirqe R. El. be divided?<sup>4</sup> Such a decision is related to the choice of representing either the units of meaning (Chapters, Paragraphs) or the physical makeup of a manuscript (Pages and Lines). As the markup will be in SGML/TEI, units of meaning might be designated by the elements

1. DIV (division) in which:
  1. The attribute TYPE would contain "chapter."
  2. The attribute ID would contain 1) an abbreviation for the name of the work (PRE), 2) the database number of the manuscript, and 3) the specific chapter number.
2. P (paragraph) in which
  1. The ID attribute would contain 1) an abbreviation for the name of the work (PRE), 2) the database number of the manuscript, and 3) the specific chapter number, and 4) the specific paragraph number.

Alternately, the physical make up of the text could be represented by the elements:

1. DIV in which:
  1. The attribute TYPE would indicate "folio."
  2. The attribute ID would contain 1) an abbreviation for the name of the work (PRE), 2) the database number of the manuscript, and 3) the specific folio number with side designated "a", "b", "c", or "d".
2. L (line) in which:
  1. The ID attribute would contain 1) an abbreviation for the name of the work (PRE), 2) the database number of the manuscript, and 3) the specific folio number with side designated "a", "b", "c", or "d" and 4) the specific line number.

Two problems emerge regarding text division, both having to do with SGML/TEI limitations and neither unique to the project. First: the problem of overlapping hierarchies. It is not presently possible to do concurrent markup, that is, to simultaneously tag material units of meaning and physical layout. Second, the TEI L tag is reserved for a line in prose not a physical line of a prose manuscript, i.e., it encloses a unit of meaning which is contained in a physical line even when the meaning may run on to the next line.

The way around this is through the use of various MILESTONE elements: MILESTONE, PB (page break), and LB (line break) which contain attributes to indicate divisions in the text, but cannot contain text.

In regard to **encoding** manuscripts of Pirqe R. El., or any rabbinic text, after long evaluation, I have concluded that the basic initial **encoding** must be in units of meaning. Rabbinic units of meaning contain quotes -- primarily from the Hebrew Bible -- which may flow through two or occasionally three lines in a manuscript. In the present state of software development, it is not possible to place an opening QUOTE tag within a line enclosed by an L tag, and then place its closing QUOTE in another line. Thus one is forced to use units of meaning to divide the text and then insert MILESTONE tags to indicate page and line breaks for each separate manuscript.

## CANONICAL REFERENCE SCHEME

One further comment regarding **encoding** Pirqe R. El. using units of meaning. The manuscripts, printed editions and electronic version of Pirqe R. El. divide the text by chapters, but do not contain paragraph divisions. Modern translations-in English, French and Spanish-do separate the text into paragraph units, but do not provide a reference scheme, beyond the page numbers of the particular translation. In sum, there is no agreed upon "canonical" reference

## system for Pirqe R. El.<sup>5</sup>

The only fully developed canonical reference system is found in the electronic text created by the "Academy of the Hebrew Language" for its "Historical Dictionary of the Hebrew Language". This electronic edition is based on one primary manuscript selected for its linguistic properties-- New York, JTS Enelow 886 (Yemen, 1654), -- corrected against four others. Consequently, it either contains material -- therefore "paragraphs" -- which are only found in manuscripts of the same family, or does not contain material -- therefore "paragraphs" -- which are found in manuscripts of a different family. Nevertheless, the AHL numbering will generally be used to establish a canonical reference system, though it may be revised as **encoding** of the different manuscripts proceeds.

## ABBREVIATIONS

As far as I can determine, this is the first SGML/TEI editing project of a medieval Hebrew work. Consequently, the issue of abbreviations and references of all kinds in the electronic context needs to be addressed. Printed editions, an especially translations of Pirqe R. El. contain notes and index references to the Bible (Hebrew Bible, LXX or NT), Apocrypha, Pseudepigrapha, the Dead Sea Scrolls, Rabbinic Literature, and the Church Fathers. Numerous modern scholarly publications (books, journals, etc.) contain references to Pirqe R. El. as well.

Several questions and issues have emerged in regard to abbreviations and references.

First, the text is in Hebrew. Consequently, when a source is originally in Hebrew, should references contain Roman Hebrew characters for titles of books?

Second, because of the differences in character representation between print media and electronic media, standard listings of abbreviations and references cannot always be used, or need to be modified. For example, references to rabbinic tractates in some abbreviation systems use scholarly transliteration. This includes diacritical marks, among which are superscript half circles to represent the Hebrew letters ALEF and AYIN at the beginning of words. Even if opening and closing parentheses are substituted for these signs, searching mechanisms don't particularly like them, or require that they be differentiated from code.

Third, the study of biblical literature is, of course, international. Western systems which provide references to biblical books are often reflective of different cultural traditions and can even differ within the same language. For example, English language countries verses from the biblical prophet Isaiah are often referenced in the following ways: Isa an (with or without a period). In German, this prophet's name is Jesaja, and referenced Jes; in French, the same prophet Esae, and referenced Es. The tendency in recent scholarly abbreviation of scriptural and related titles is not to include period after the book reference. Thus, Isa 1:5. The space after the name and semi-colon between the chapter and verse work well in printing, but not in an electronic context.

Finally, even in so-called standard works, such as the Bible, differences exist in verse numbering (i.e., various editions of the HB, NT or LXX).<sup>6</sup> Thus, it becomes necessary to indicate the specific edition of the work in a bibliographical note.

How does one proceed without reinventing the wheel? First, by choosing existing standards and indicating where modification is necessary. Second, by exploiting the advantages of electronic search mechanisms, one of which is to use the period as a delimiter, setting off parts of a reference.

The language for the scholarly notations and tagging of this project will be English, and the reference standard that of the American Academy of Religion/Society of Biblical Literature, as found in *SBL: Membership Directory and Handbook*, 1994, pp. 224-240.<sup>7</sup> However, superscript for "ALEF" and "AYIN" as well as other diacritic marks for rabbinic texts are omitted. Where the reference contains two words, no space should be placed between the words; ex. "Ros Has" =<Rosh Hashanah> would appear "RosHas." If at all possible, each source reference should be composed of four parts, each part separated by a period; ex. "HB.Gen.20.15."<sup>8</sup> The first part represents the general body of

literature (HB=Hebrew Bible), the second the specific text (Gen=Genesis), the third either the chapter (20=chapter 2 or the folio, the fourth either the verse (15=verse 15) or the column.

### Reference examples:

- HB.Gen.20.15=Hebrew Bible, Genesis 20:15.
- LXX.Gen.20.15=Septuagint, Genesis 20:15.
- NT.Matt.5.6=(Greek)New Testament, Matthew, 5:6.
- Rab.mAbot.1.3=Rabbinic Literature, Mishna, Abot 1:3.
- Rab.bBer.25.a= Rabbinic text, Babylonian Talmud, Berakot 25a.
- QL.1QapGen.1.3=Qumran Literature, Genesis Apocryphon from Qumran Cave 1.

Note that there should be a period even prior to the page or folio in a Talmud reference.

Such references are to be encoded as text (CDATA) as in the following example:<sup>9</sup>

< CIT > < BIBL > < ADD > HB.Gen.1.1 < /ADD > < /BIBL > < Q > BRAWYT BRA ALHYm < /Q > < /CIT >.

This form of **encoding** canonical references in the text itself allows the following to happen:

Searches can be created for

- The body of literature (HB)
- The body of literature (HB), and the specific biblical book (Gen).
- The body of literature (HB), the specific biblical book (Gen) and the chapter (1).
- the body of literature (HB), the specific biblical book (Gen) the chapter (1) and the verse.

In addition:

- The entire canonical reference enclosed in >< BIBL > < ADD > HB.Gen.1.1 < /ADD > < /BIBL > can be hidden or visible.
- The responsibility for the identification of the reference can be designated in the RESP attribute of the element ADD.
- It would be easy to replace the English reference scheme for biblical or rabbinic books and with proper software reverse the visual presentation of the reference, with numbers retained in proper order (HB.Gen.1.1 -> 1.1.AR.B.KNT).

## CONCLUSIONS

SGML/TEI markup is particularly useful for scripturally based text, i.e., texts from the vast literatures of Judaism, Christianity and Islam which frequently cite biblical or koranic verses. There are numerous genres in these religious literatures (exegetical works, homilies, scriptural essays, dialogues, legal texts, liturgical texts, religious poetry, etc.) They all have in common the citation of texts sacred to a religious community, the frequent mention of characters, places and institutions found in such texts, plus references to later individuals, places and institutions. In addition, the texts are often *macaronic*, i.e., they contain more than one human language.

Such texts offer particular problems for electronic presentation, apart from the issues of the non-existence of SGML software for viewing correctly encoded Near-Eastern languages. This paper has focused on technical issues, the solution of which will be indicated in an **Encoding Manual** used both as a supplement to viewing the electronic text and as a guide for those participating in the **encoding** process.

## ENDNOTES:

(An earlier version of this paper was presented at the ALLC-ACH 1996 Joint International Conference, Bergen, Norway, June 1996.)

1. See the numerous articles cited in notes by Jacob Elbaum, "Rhetoric, Motif and Subject-Matter: Toward an Analysis of Narrative Technique in Pirke de-Rabbi Eliezer," *Jerusalem Studies in Jewish Folklore*, XIII-XIV, (1992), 99-126. Pirqe R. El. was translated into Latin by the sixteenth century: R. Eliezer f. Hircani: *Liber sententiarum Judiacarum*, trans. Konrad Pellikan, (1546) [see comment on this by Hans Jakob Haag, *Pirqe DeRabbi Eli'ezer Kap. 43*, Magisterarbeit, Köln, 1978. In the twentieth century Pirqe R. Eliezer has been translated into English, French and Spanish: *Pirké De Rabbi Eliezer*, trans. Gerald Friedlander, (London, 1916; reprint, New York: Hermon Press, 1965) *Pirqué De Rabbi 'Eliezer: Leçons De Rabbi Eliezer*, trans. Marc-Alain Ouaknin, Eric Smilevitch and Pierre-Henri Salfati. Paris, (Verdier, 1984); and *Los Capítulos De Rabbi Eliezer*, trans. M. Pérez Fernández, (Valencia, 1984).

2. Bar Ilan Database (Responsa Database, Bar Ilan University); STM Database (Polytext, Jerusalem) and Davka database of Rabbinic Literature.

3. For the debate on this issue between Schäfer and Milikowsky, see: Milikowsky, Chaim, "The Status Quaestionis of Research in Rabbinic Literature." *JJS* 39, no. 2 (1988): 201-211 and Schäfer, Peter. "Once Again the Status Quaestio of Research in Rabbinic Literature: An Answer to Chaim Milikowsky." *JJS* 40, no. 1 (1989): 89-94. In addition, Malachi Beit-Arié has approached the same question from the perspective of codicological issues. See: Malachi Beit Arié, "Transmission de textes par scribes et copistes. Interférences inconscientes et critiques", *Les problèmes posés par l'édition critique des textes anciens et médiévaux* ('Louvain-la-Neuve, 1992), 175.

4. Peter Robinson has written, "perhaps the most important decision an encoder of scholarly text must face is how the text should be divided (Transcription, p. 64)."

5. Traditional citing most often utilizes the pagination of the edition of the RaDaL, the page division of the edition of Higger, or occasionally reference to the "critical edition" of C. M. Horowitz. The problems of all these texts will be discussed in a separate document "Introduction: the Need for a Critical Edition of Pirqe R. El."

6. My thanks to Robin Cover for reminding me of this.

7. For abbreviations of journals, etc., additional items are found in the Index of Articles on Jewish Studies, (The Jewish National and University Library: Jerusalem, 1995 and earlier), "List of Periodicals and the Collections and their Abbreviations," and International Glossary of Abbreviations for Theology and Related Subjects, ed. Sigfried Schwertner (Walter de Gruyter: Berlin and New York, 1974).

8. The MILESTONE tag LB (line break) will also use a four part structure for the ID attribute. Example: PRE.04.26b This refers to the work Pirq. R. El.; manuscript 04 (so designated in manuscript database0); folio 26b (+ a = rechto o = verso); line 1.

9. My thanks to Lou Burnard for suggesting this tagging and **encoding** for canonical references. In the published abstract of this paper I had stated: "Such references are to be used in the attribute "N" for QUOTE and in various notation and bibliographical elements." It soon became clear to me that it is not possible to do the kind of searching indicated above if the canonical reference was placed within an element attribute. Once I had utilized this CIT scheme the additional advantages for visibility, identification of responsibility and ease of revision became clear.

## BRIEF BIBLIOGRAPHY ON ELECTRONIC TEXT ENCODING AND ENCODING OF HEBREW TEXTS

1. Nancy Ide and Jean Veronis, eds., **Text Encoding** Initiative: Background and Context (Kluwer Academic Publish Dordrecht/Boston/London, 1995).
2. Stephen A. Kaufman, The Comprehensive Aramaic Lexicon: Text Entry and Format **Manual**, Comprehensive Aramaic Lexicon, Baltimore 1987.
3. Chaim Milikowsky, The Henkind Talmud Text Database, Lieberman Institute for Talmudic Research (JTS): "Directions for Text Copyists, Hebrew and English versions," Jerusalem and New York, no date.
4. Leib Moscovitz, Responsa Version 3.0: User's Guide, Bar-Ilan University, Ramat Gan 1994.
5. Peter Robinson, The Transcription of Primary Textual Sources Using SGML, Office for Humanities Communicat Publications Number 6, Oxford 1994.
6. Peter Schäfer and/or Gottfried Reeg, "Konventionen zur Aufnahmen von Handschriften für die Datenverarbeitung Berlin, Stand: November 1991.
7. Sperberg-McQueen and Lou Burnard, ed. Guidelines for Electronic Text **Encoding** and Interchange (TEI3), Chicago, Oxford, April 8, 1994.

---

## ELECTRONIC FILE INFORMATION

- ELECTRONIC TITLE: Pirqe R. El.: Creating an **Encoding Manual**
- TITLE: Electronic Edition of the Midrash Pirqe Rabbi Eliezer: Creating an **Encoding Manual**
- KEYWORDS: Midrash, **Manual**, Hebrew, SGML, TEI
- AUTHOR: Lewis M. Barth
- AFFILIATION: Hebrew Union College - Los Angeles

---

*Lewis M. Barth*  
*Hebrew Union College - Jewish Institute of Religion*  
*3077 University Avenue*  
*Los Angeles, California 90007-3796*  
*Phones:*  
*Office: (213) 749-3424*  
*Office FAX: (213) 749-1192*  
*You can e-mail me at: [lbarth@bcf.usc.edu](mailto:lbarth@bcf.usc.edu)*